generating a score for at least a subset of said entries of said matrix, such that a score reflects a relative theme pair strength between two themes represented by said entry for said documents.

13. (Once Amended) The computer readable medium as set forth in claim 12, wherein the step of generating a score for at least a subset of said entries of said matrix comprises the steps of:

a4

calculating a plurality of products for an entry by multiplying theme strengths corresponding to two themes represented by said entry for each document that includes said two themes represented by said entry; and

summing said products for an entry to generate said score.

## REMARKS

Reconsideration of the application in view of the above amendments and the following remarks is respectfully requested.

Overview of the Claimed Invention:

A system generates cross-references among categories in a knowledge base. A plurality of themes are extracted from a corpus of documents. A theme identifies subject matter contained in a corresponding document. The system generates a plurality of scores such that each score identifies a relative theme pair strength between theme pairs. In

general, a theme strength reflects the amount of subject matter contained in a document for a corresponding theme relative to other themes in the document. Thereafter, the most related theme pairs are selected as indicated by the scores. Category pairs of the knowledge base are then selected by mapping the themes of the selected theme pairs to corresponding categories of the knowledge base. A cross reference between categories of the category pairs in the knowledge base are generated so as to identify an association between the category pairs.

**Rejection of the Claims Under 35 U.S.C. § 102**

In the Office Action dated October 3, 2002, claims 1-15 were rejected under 35 U.S.C. § 102 as being anticipated by US Patent 5,953,718, issued to Wical (hereafter referred to as *"Wical"*). Also, the abstract of the disclosure was objected to for containing the phrase "is disclosed."

Overview of *Wical*:

*Wical* discloses a research mode for a search and retrieval system. In general, the search and retrieval system infers an answer to a user query. The search and retrieval system stores point of view gists for documents. A point of view gist provides a synopsis for a document with a slant toward a particular topic or theme. First, the user query is processed, including identifying a query term strength based on the content carrying strength of the corresponding word. Table 2 of *Wical* shows example query terms and

corresponding query terms strengths. The system identifies documents relative to the query, as well as the themes and corresponding theme weights for those documents. Table 3 of *Wical* lists example documents along with their themes and theme weights. A total score is generated for each document. Table 4 shows total scores for each document in the example.

The research document processing builds research documents by answering a response to the query from more than one document. The research document processing selects, starting from the most relevant document, additional documents that have a common theme as well as themes required to satisfy the remaining portions of the query. (Col 11, lines 60 - 63). Then, point of view gists, with slants toward the corresponding themes, are selected to compile the research document. (Col 12, lines 15-17).

**A. The Cited References Do Not Disclose An Automated Process To Generate Cross-References Between Categories In A Knowledge Base.**

A lexicon consists of a very large repository of language specific words/phrases, their corresponding parts of speech information, and the relationships to each other. Typically, the entries in lexicons are arranged in a tree shaped hierarchy. In addition, other relationships establish associations between two words in the lexicons. This type of relationship is referred to as a "cross reference relationship." (Specification, page 2, lines 18-22). These lexicons are mostly language dependent and are manually constructed. A typical lexicon contains about half a million words/phrases for the English language. The

process to manually establish relationships between such a large number of words is very time consuming. (Specification, page 2, lines 10-17).

The claimed invention recites elements "for generating cross-references among categories in a knowledge base." For example, the last element of claim 1 sets forth:

*generating a cross reference in said knowledge base between categories of said category pairs, wherein said cross reference identifies an association between said category pairs.*

As such, the claimed invention sets forth a process to automatically establish cross-reference relationships in a knowledge base using a repository of documents. Applicant respectfully contends that *Wical* does not disclose or suggest a process to automatically establish cross-reference relationships in a knowledge base.

The claimed invention also solves the problem of how to establish cross reference relationships with words not currently present in the lexicon, even though these relationships are pertinent to the documents under analysis. In addition, the claimed invention solves the problem of how to establish new cross references within existing words based on the specific use of each of the words in the data set under analysis. (Specification, page 3, lines 7-14). Applicant respectfully contents that the cited references to not disclose a process that automatically generates cross references between

categories of a knowledge base based on a corpus of documents.

The Office Action references *Wical* at column 16, lines 54-57 in support of the rejection of the "generating a cross reference in said knowledge base between categories of said category pairs" limitation. At this location, *Wical* discloses a knowledge base "augmented with additional terminology including cross references and links, terminology/categories." Thus, *Wical* does disclose a knowledge base with cross references. However, *Wical* does not disclose or suggest a teaching to automatically generate these cross-references for a knowledge base. The teachings of the present invention are applicable to the knowledge base of *Wical*, such that such a knowledge base may be automatically expanded through use of the claimed invention.

**B. The Cited References Do Not Generate Scores To Identify Theme Pair Strengths For Pairs Of Themes.**

Amended claim 1 recites:

*generating a theme strength for said themes, said theme strength reflects the amount of subject matter contained in a document for a corresponding theme relative to other themes in said document;*

*generating a plurality of scores, from said theme strengths, to identify a relative theme pair strength for at least one pair of said themes extracted from said documents;*

Thus, as claimed, the process generates theme strengths for themes, and generates scores to reflect a theme pair strength for *pairs of themes*. Applicant respectfully contends that *Wical* does not disclose or suggest a system that generates scores to reflect a strength for a pair of themes.

> *Wical* discloses generation of the score as follows:
>
> The first column labeled "documents", identifies each document in the example documents set. The second column, labeled "#of query terms", is generated by multiplying the number of query terms the documents satisfy times a factor (e.g., $1x10^6$). (Col. 11, lines 15-19).

*Wical* discloses the process of generating the total query weight as:

> The third column of Table 4 in *Wical* is generated by multiplying the total query weight by the factor of 1000. (Col. 11, lines 23-24).

The third column of Table 4 lists the total theme weights for the corresponding document. (Col. 11, lines 27-28). The total score is calculated as follows.

> The fourth column, labeled "total score", is calculated by summing the values from columns 2, 3, and 4. In general, the total score provides a quantitative value to measure the relevance of the corresponding document to the query terms. (Col. 11 lines 31-35).

Accordingly, *Wical* discloses that the total score measures the relevance of the corresponding document to the query terms. In contrast, the "scores" of the claimed invention identify relative theme strengths for theme pairs. As such, the amended claims are patentable over the cited reference.

## CONCLUSION

In view of the foregoing, it is submitted that the claims are in condition for allowance. Reconsideration of the rejections and objections is requested. Allowance is earnestly solicited at the earliest possible date.

Respectfully submitted,

STATTLER JOHANSEN & ADELI LLP

Dated: 2/3/2003

John Stattler
Reg. No. 36,285

Stattler, Johansen & Adeli LLP
PO Box 51860
Palo Alto, CA 94303-0728
Phone: (650) 934-0470 ext.100
Fax: (650) 934-0475

## The Amended Claims

The following pages provide the amended claims with the amendments marked with deleted material in [brackets] and new material underlined to show the changes made.

1. (Once Amended) A method for generating cross-references among categories in a knowledge base, said method comprising the steps of:

extracting, from a plurality of documents, a plurality of themes, wherein a theme identifies subject matter contained in a corresponding document;

generating a theme strength for said themes, said theme strength reflects the amount of subject matter contained in a document for a corresponding theme relative to other themes in said document;

generating a plurality of scores, from said theme strengths, [such that each score] to identify[ies] a relative theme pair strength [among] for at least one pair [theme pairs] of said themes extracted from said documents[, said theme strength reflects the amount of subject matter contained in a document for a corresponding theme relative to other themes in said document];

selecting theme pairs based on said scores;

selecting category pairs in said knowledge base by mapping said themes of said theme pairs selected to corresponding categories of said knowledge base; and

generating a cross reference in said knowledge base between categories of said category pairs, wherein said cross reference identifies an association between said category pairs.

2. (Once Amended) The method as set forth in claim 1, wherein the step of generating a plurality of scores [comprising] comprises the steps of:

generating a matrix comprising a plurality of columns and rows to form a plurality of entries, wherein each column represents one of said themes and each row represents one of said themes; and

generating a score for at least a subset of said entries of said matrix, such that a score reflects a relative theme pair strength between two themes represented by said entry for said documents.

3. (Once Amended) The method as set forth in claim 2, wherein[:

the step of extracting a plurality of themes further comprises the step of generating theme strengths for each theme extracted; and]

the step of generating a score for at least a subset of said entries of said matrix comprises the steps of:

calculating a plurality of products for an entry by multiplying theme strengths corresponding to two themes represented by said entry for each document that includes said two themes represented by said entry; and

summing said products for an entry to generate said score.

37. (Once Amended) A system comprising:

search and retrieval module for receiving a user query and for generating a query response including query feedback;

a knowledge base, coupled to said search and retrieval module, for storing relationships among terminology for use as query feedback;

a knowledge base processing system, coupled to said knowledge base for processing a plurality of documents and automatically extending said relationships among said terminology in said knowledge base, said knowledge base processing system for extracting, from said documents, a plurality of themes, wherein a theme identifies subject matter contained in a corresponding document, for generating a theme strength for said themes, said theme strength reflects the amount of subject matter contained in a document for a corresponding theme relative to other themes in said document, for generating a plurality of scores, from said theme strengths, [such that each score] to identify[ies] a relative theme pair strength [among] for at least one pair [theme pairs] of said themes extracted from said documents[, said theme strength reflects the amount of subject matter contained in a document for a corresponding theme relative to other themes in said document], for selecting theme pairs based on said scores, for selecting category pairs in

said knowledge base by mapping said themes of said theme pairs selected to corresponding categories of said knowledge base, and for generating a cross reference in said knowledge base between categories of said category pairs, wherein said cross reference identifies an association between said category pairs.

7. (Once Amended) The system as set forth in claim 6, wherein the knowledge base processing system further for generating a matrix comprising a plurality of columns and rows to form a plurality of entries, wherein each column represents one of said themes and each row represents one of said themes and for generating a score for at least a subset of said entries of said matrix, such that a score reflects a relative theme pair strength between two themes represented by said entry for said documents.

8. (Once Amended) The system as set forth in claim 7, wherein the knowledge base processing system further [for generating theme strengths for each theme extracted] for calculating a plurality of products for an entry by multiplying theme strengths corresponding to two themes represented by said entry for each document that includes said two themes represented by said entry, and for summing said products for an entry to generate said score.

11. (Once Amended) A computer readable medium comprising a plurality of instructions, which when executed, causes the computer to perform the steps of:

extracting, from a plurality of documents, a plurality of themes, wherein a theme identifies subject matter contained in a corresponding document;

generating a theme strength for said themes, said theme strength reflects the amount of subject matter contained in a document for a corresponding theme relative to other themes in said document;

generating a plurality of scores, from said theme strengths, [such that each score] to identify[ies] a relative theme pair strength [among] for at least one pair [theme pairs] of said themes extracted from said documents[, said theme strength reflects the amount of subject matter contained in a document for a corresponding theme relative to other themes in said document];

selecting theme pairs based on said scores;

selecting category pairs in said knowledge base by mapping said themes of said theme pairs selected to corresponding categories of said knowledge base; and

generating a cross reference in said knowledge base between categories of said category pairs, wherein said cross reference identifies an association between said category pairs.

12. (Once Amended) The computer readable medium as set forth in claim 11, wherein the step of generating a plurality of scores [comprising] comprises the steps of:

generating a matrix comprising a plurality of columns and rows to form a plurality of entries, wherein each column represents one of said themes and each row represents one of said themes; and

generating a score for at least a subset of said entries of said matrix, such that a score reflects a relative theme pair strength between two themes represented by said entry for said documents.

13. (Once Amended) The computer readable medium as set forth in claim 12, wherein[:

the step of extracting a plurality of themes further comprises the step of generating theme strengths for each theme extracted; and]

the step of generating a score for at least a subset of said entries of said matrix comprises the steps of:

calculating a plurality of products for an entry by multiplying theme strengths corresponding to two themes represented by said entry for each document that includes said two themes represented by said entry; and

summing said products for an entry to generate said score.